Amendments to the Drawings:

The attached sheet of drawings includes changes to the drawings. This sheet, which includes Figs. 2, 7, and 8, replaces the original sheet including Figs. 2, 7, and 8. In Fig. 2, the word "CONDUOR" has been corrected to "CONTOUR". In Fig. 7, the word "TABLEL" has been corrected to "TABLE". In Fig. 8, the word "CONDUOR" has been corrected to "CONTOUR".

REMARKS

The specification has been reviewed, and clerical errors are corrected.

In paragraph 1 of the Action, the drawings were objected to because of the informalities. In view of the objection, the sheet of the corrected drawings including Figs. 2, 7, and 8 has been filed.

In paragraph 2 of the Action, claim 3 was objected to because of the informalities. In view of the objection, claim 3 has been amended to correct the informalities.

In paragraph 4 of the Action, claim 8 was rejected under the second paragraph of 35 U.S.C. 112. In view of the rejection, claim 7 has been amended to alleviate the rejection.

In paragraph 6 of the Action, claims 1 and 2 were rejected under 35 U.S.C. 103(a) being unpatentable over applicant's admitted prior art, in view of *Otsuka* (US Patent No. 6,546,367).

In paragraph 7 of the Action, claims 3 and 4 were rejected under 35 U.S.C. 103(a) being unpatentable over applicant's admitted prior art, in view of *Otsuka*, further in view of *Vermeulen et al.* (US Patent No. 6,810,379).

In paragraph 8 of the Action, claims 5 and 6 were rejected under 35 U.S.C. 103(a) being unpatentable over applicant's admitted prior art, in view of *Hara et al.* (US Patent No. 5,615,300).

In paragraph 9 of the Action, claims 7 to 9 were rejected under 35 U.S.C. 103(a) being unpatentable over applicant's admitted prior art, in view of *Rye* (*Speech Synthesis at Higher Speaking Rates*).

In paragraph 10 of the Action, claims 10 and 11 were rejected under 35 U.S.C. 103(a) being unpatentable over applicant's admitted prior art.

In paragraph 11 of the Action, claims 12 to 14 were rejected under 35 U.S.C. 103(a) being unpatentable over applicant's admitted prior art, in view of *Walsh* (US Patent Application Publication 2003/0014253).

The Applicants respectfully traverse the rejections and request reconsideration. In view of the rejections cited in paragraphs 6 and 7, claims 1 and 3 have been amended to clarify the features of the invention. With the amendments, claims 1 to 4 are not unpatentable over applicant's admitted prior art, in view of the cited references, for the reasons explained below. Claims 5 to 14 are not unpatentable over applicant's admitted prior art, in view of the cited references, for the reasons explained below.

As recited in claim 1, a method of the invention controls high-speed reading in a text-to-speech conversion system. The text-to-speech conversion system includes a text analysis module for generating a phoneme and prosody character string from an input text; a prosody generation module for generating a synthesis parameter of at least a voice segment, a phoneme duration, and a fundamental frequency for the phoneme and prosody character string; a voice segment dictionary in which voice segments as a source of voice are registered; and a speech generation module for generating a synthetic waveform by waveform superimposition by referring to the voice segment dictionary.

In particular, the method comprises the step of providing the prosody generation module with a phoneme duration determination unit that includes both a duration rule table containing

empirically found phoneme durations and a duration prediction table containing phoneme durations predicted by statistical analysis. The phoneme duration determination unit determines a phoneme duration by using the duration rule table when a user-designated utterance speed exceeds a threshold contained in the duration rule table, and by using the duration prediction table when the utterance speed does not exceed the threshold. That is, the phoneme durations in the duration rule table are empirically found in advance, and the threshold is one of the phoneme durations in the duration rule table. The phoneme duration determination unit selects the method of determining the phoneme duration based on whether the utterance speed exceeds the threshold contained in the duration rule table.

Otsuka discloses a speech synthesizing method and apparatus as well as a storage medium for setting a phoneme duration for a phoneme string to achieve a specified speech-production time and provide a natural phoneme duration regardless of a length of speech production time. In *Otsuka*, Fig. 2 shows a block diagram of a flow structure of the speech synthesizing apparatus. In Fig. 2, a phoneme duration setting unit 5 sets a phoneme duration in accordance with control data, representing speech production speed stored in a control data storage unit 2. According to *Otsuka*, using the phoneme duration value, the phoneme duration is determined according to the equation (3a). When the obtained phoneme duration is smaller than a threshold value, the phoneme duration is determined according to the equation (3b), in which the phoneme duration is equal to the threshold value, so that reproduced speech becomes natural (col. 3, line 16 to col. 4, line 60).

In the invention recited in claim 1, the phoneme duration determination unit determines the phoneme duration by using the duration rule table when the utterance speed exceeds the threshold contained in the duration rule table, and by using the duration prediction table when the utterance speed does not exceed the threshold. The phoneme durations in the duration rule table are empirically found in advance, and the threshold is one of the phoneme durations in the duration rule table. The phoneme duration determination unit selects the method of determining the phoneme duration based on whether the utterance speed exceeds the threshold contained in the duration rule table.

On the other hand, *Otsuka* fails to elaborate the nature of the threshold value. In Otsuka, it is simply stated that when the obtained phoneme duration is smaller than the threshold value, the phoneme duration is determined according to the equation (3b), in which the phoneme duration is equal to the threshold value. There is no disclosure regarding the threshold empirically found in advance. Accordingly, the threshold value disclosed in *Otsuka* is totally different from the threshold claimed in the invention.

Further, in the invention, the duration prediction table is used to set the phoneme duration when the utterance speed does not exceed the threshold. On the other hand, in *Otsuka*, the phoneme duration is equal to the threshold value when the phoneme duration is smaller than the threshold value. Accordingly, the method of setting the phoneme duration claimed in the invention is totally opposite to that disclosed in *Otsuka*.

Therefore, *Otsuka* does not disclose nor suggest the features of the invention recited in claim 1. Further, even though *Otsuka* is combined with Applicant's admitted prior art, the invention recited in claim 1 is not obvious.

As recited in claim 3, a method of the invention controls high-speed reading in a text-to-speech conversion system. The text-to-speech conversion system includes a text analysis module for generating a phoneme and prosody character string from an input text; a prosody generation module for generating a synthesis parameter of at least a voice segment, a phoneme duration, and a fundamental frequency for the phoneme and prosody character string; a voice segment dictionary in which voice segments as a source of voice are registered; and a speech generation module for generating a synthetic waveform by waveform superimposition while referring to the voice segment dictionary.

In particular, the method comprises the step of providing the prosody generation module with a pitch contour determination unit that has both an empirically found rule table and a prediction table predicted by statistical analysis. The pitch contour determination unit determines a pitch contour by determining both accent and phrase components with the rule table when a user-designated utterance speed exceeds a threshold contained in the rule table, and with the prediction table when the utterance speed does not exceed the threshold.

As explained above, in Otsuka, using the phoneme duration value, the phoneme duration is determined according to the equation (3a). When the obtained phoneme duration is smaller than the threshold value, the phoneme duration is determined according to the equation (3b), in which the phoneme duration is equal to the threshold value. There is no disclosure regarding the threshold empirically found in advance. Accordingly, the threshold value disclosed in *Otsuka* is totally different from the threshold claimed in the invention.

Vermeulen et al. has disclosed a client/server architecture for text-to-speech synthesis. In Fig. 1 in *Vermeulen et al.*, a text-to-speech system 10 is provided with a prosody generation unit 16. The prosody generation unit 16 produces timing and pitch information for speech synthesis. According to *Vermeulen et al.*, the pitch is determined from a rule set or statistical model (col. 2, line 1 to line 21). In the invention, the pitch contour determination unit determines a pitch contour by determining both accent and phrase components with the rule table when a user-designated utterance speed exceeds a threshold contained in the rule table, and with the prediction table when the utterance speed does not exceed the threshold. In *Vermeulen et al.*, it is simply stated that the pitch is determined from a rule set or statistical model. There is no disclosure or suggestion regarding the method of setting the pitch contour base on the threshold as claimed in the invention.

Therefore, neither *Otsuka* nor *Vermeulen et al.* discloses or suggest all the features of the invention. Even though *Otsuka* and *Vermeulen et al.* are combined with Applicant's admitted prior art, the invention is not obvious.

As recited in claim 5, a method of the invention controls high-speed reading in a text-to-speech conversion system. The text-to-speech conversion system includes a text analysis module for generating a phoneme and prosody character string from an input text; a prosody generation module for generating a synthesis parameter of at least a voice segment, a phoneme duration, and a fundamental frequency for the phoneme and prosody character string; a voice segment dictionary in which voice segments as a source of voice are registered; and a speech generation module for generating

a synthetic waveform by waveform superimposition by referring to said voice segment dictionary.

In particular, the method comprises the step of providing the prosody generation module with a sound quality coefficient determination unit that has a sound quality conversion coefficient table for changing the voice segment to switch sound quality. When a user-designated utterance speed exceeds a threshold, the sound quality coefficient determination unit selects a coefficient from the sound quality conversion coefficient table such that sound quality does not change.

Hara has disclosed a method of generating synthesized speech while allowing a period of time required for speech synthesis and the quality of synthesized speech to be varied by varying the order of filtering for speech synthesis (col. 2, line 37 to line 42). In Fig. 3 in *Hara*, a speech synthesizing apparatus has a speech synthesizer 16 for generating a sound source based on phonetic parameters generated by a synthetic parameter generator 15. The speech synthesizer 16 also effects filtering on the sound source with a filter according to order information supplied from a mode selector 21. The mode selector 21 selects one of the order or arrangement information supplied from an input unit 11 and a rate controller 20 based on mode selecting information stored in a rate information file 18 (col. 11 line 20 to col. 12, line 3). The rate information file 18 stores data about average processing rate required to generate real-time synthesized speech depending on phonetic parameter order. When the speech synthesizing process is carried out with a specific phonetic parameter order, an activity ratio of CPU has an upper limit. In other words, the mode selector 21 selects one of the order or arrange information based on the activity ratio of CPU.

In the invention recited in claim 5, when the user-designated utterance speed exceeds the threshold, the sound quality coefficient determination unit selects the coefficient from the sound quality conversion coefficient table such that sound quality does not change. In other words, the sound quality coefficient determination unit selects the coefficient based on the utterance speed. In *Hara*, the mode selector selects one of the order or arrange information based on the activity ratio of CPU. The activity ratio of CPU is affected by various tasks other than the utterance speed performed by the CPU, and is not equivalent to the utterance speed. Therefore, the method disclosed in *Hara* is different from the method claimed in the invention recited in claim 5. Even though *Hara* is combined with Applicant's admitted prior art, the invention recited in claim 5 is not obvious.

As recited in claim 7, a method of the invention controls high-speed reading in a text-to-speech conversion system. The text-to-speech conversion system includes a text analysis module for generating a phoneme and prosody character string from an input text; a prosody generation module for generating a synthesis parameter of at least a voice segment, phoneme duration, and fundamental frequency for the phoneme and prosody character string; a voice segment dictionary in which voice segments as a source of voice are registered; and a speech generation module for generating a synthetic waveform by waveform superimposition by referring to the voice segment dictionary.

In particular, the method comprises the step of providing the prosody generation module with both a pitch contour correction unit for outputting a pitch contour corrected according to an intonation level designated by the user and a switch for determining whether a

base pitch is added to the pitch contour corrected according to the user-designated utterance speed. When the utterance speed exceeds a threshold, the switch is controlled not to change the base pitch.

Rye has discussed speech synthesis at higher speaking rates. *Rye* has stated that at very low pitch the voice pulses are relatively far apart in time, consequently they do not sample synthesizer phonetic segments very often. The perception of short voiced sounds in which the vocal tract parameters, for instance formant frequencies, are varying most rapidly, may then be impaired. Conversely, too high a pitch value may produce voice harmonics too far apart to sample the synthesized vocal tract resonances effectively, resulting again in a loss in intelligibility.

However, *Rye* fails to mention specifically how to adjust the base pitch, mere mentioned that the voice pitch affects intelligibility at high speaking rate. In the invention recited in claim 7, when the utterance speed exceeds the threshold, the switch is specifically controlled not to change the base pitch. Therefore, *Rye* does not disclose the features of the invention recited in claim 7. Further, even though *Rye* is combined with Applicant's admitted prior art, the invention recited in claim 7 is not obvious.

As recited in claim 10, a method of the invention controls high-speed reading in a text-to-speech conversion system. The text-to-speech conversion system includes a text analysis module for generating a phoneme and prosody character string from an input text; a prosody generation module for generating a synthesis parameter of at least a voice segment, a phoneme duration, and a fundamental frequency for said phoneme and prosody character string; a voice segment dictionary in which voice segments as a source of voice are registered; and a speech generation module for

generating a synthetic waveform by waveform superimposition while referring to said voice segment dictionary.

In particular, the method comprises the step of providing the speech generation module with signal sound generation means for inserting a signal sound between sentences to indicate an end of a sentence when a user-designated utterance speed exceeds a threshold.

In the Action, the examiner admitted that the applicant's admitted prior art discloses all of the features except the step of providing the speech generation module with signal sound generation means. The examiner further stated that indexing spoken speech with signal sounds helps a listener to easily understand. However, the examiner asserted general knowledge without showing any concrete reference. In is well known standard that patent examiners carry the burden of establishing a *prima facie* case of obviousness by showing a reference.

In the invention, the speech generation module is provided with the signal sound generation means for inserting a signal sound between sentences. A structural configuration of the signal sound generation means is described in detail from line 27 on page 40 of the specification, and is shown in Fig. 12. The examiner fails to show any reference disclosing the specific configuration of the signal sound generation means disclosed in the specification. Therefore, the invention recited in claim 10 is not obvious over the applicant's admitted prior art.

As recited in claim 12, a method of the invention controls high-speed reading in a text-to-speech conversion system. The text-to-speech conversion system includes a text analysis module for generating a phoneme and prosody character string from an input text; a prosody generation module for generating a synthesis

parameter of at least a voice segment, a phoneme duration, and a fundamental frequency for the phoneme and prosody character string; a voice segment dictionary in which voice segments as a source of voice are registered; and a speech generation module for generating a synthetic waveform by waveform superimposition by referring to the voice segment dictionary.

In particular, the method comprises the step of providing the prosody generation module with a phoneme duration determination unit for performing a process in which when a user-designated utterance speed exceeds a threshold, an utterance speed of at least a leading word in a sentence is returned to a normal utterance speed.

Walsh has disclosed a method and device for converting text to speech such that playing duration is decreased without significantly reducing the comprehensibility of the generated speech. Figs. 7A and 7B in *Walsh* show the paying of a text segment "The motorcycle is in the garage" with and without acceleration in accordance with technology *Walsh* developed. In the playing, the keyword "garage" has been maintained at its default rate. Note that the word "garage" is located at the end of the text segment.

In the invention recited in claim 12, when the utterance speed exceeds the threshold, the phoneme duration determination unit performing a process in which the utterance speed of a leading word in a sentence is returned to a normal utterance speed. In *Walsh*, the keyword, not the leading word of a sentence as claimed in the invention, is maintained at its default rate. In *Walsh*, there is no disclosure regarding the method in which the leading word is returned to a normal utterance speed according to the utterance speed as claimed in the invention. Therefore, *Walsh* does not

disclose or suggest the features of the invention recited in claim 12. Further, even though *Walsh* is combined with Applicant's admitted prior art, the invention recited in claim 12 is not obvious.

As explained above, the cited references do not disclose or suggest all of the features of the invention recited in claims 1, 3, 5, 10 and 12. Further, even though the cited references are combined with Applicant's admitted prior art, the invention is not obvious. Therefore, the invention is not patentable over the applicant's admitted prior art in view of the cited references.

Reconsideration and allowance are earnestly solicited.

Three-month extension of time is requested. The credit card payment form in the amount of $1,020 has been attached herewith.

Respectfully submitted,

Kazunao Kubotera
Reg. No. 51,194

TAKEUCHI & KUBOTERA, LLP
200 Daingerfield Rd.
Suite 202
Alexandria, VA 22314
Tel. (703) 684-9777
Fax. (703) 684-1390